PRISON SERVICE JOURNAL

March 2017 No 230

The predictive validity of the LSI-R, the VRAG, and the PCL-R for prison misconduct among lifetime prisoners

***Dr Ulrika Haggård**, is based at Swedish National Board of Forensic Medicine and Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden. **Dr Åsa Eriksson** is based at the Department of Clinical Neuroscience, Division of Psychology, Karolinska Institutet, Stockholm, Sweden.*

Background

During the last decade, there has been an increased concern of security-related issues in the Swedish prison system, including a focus on prison misconduct, in particular major misconduct such as violence. To facilitate allocation of resources, there is a need of risk assessment instruments to correctly identify offenders with the highest likelihood of committing major misconduct. The present study explores risk assessment instruments of prison misconduct in a total cohort of lifetime prisoners in Sweden.

Risk assessment instruments

There is no such thing as a perfect risk assessment instrument. Deciding which instrument to use is a balance of pros and cons in relation to the population, setting and purpose of the assessment.[1] Not surprisingly, there is consensus among researchers that risk assessment instruments should be high in predictive accuracy. Based on a meta-analysis, Singh and coworkers[2] concluded that the best predictive accuracy is acquired when the instrument is based on a population with similar demographic features as the one of interest. Haggård-Grann[3] recommended that risk assessments in clinical settings, conducted with the purpose to facilitate risk management, should include dynamic and changeable risk factors. In his oft-cited article on the guidelines of the selection and use of risk assessment instruments, Bonta[4] recommended that risk assessment instruments should derive from relevant theory and include several areas of interest. There are also practical considerations such as the cost and ease of the instrument.[5] Campbell and coworkers[1] counselled researchers not to develop new scales but to validate the existing ones.

To current knowledge, there are only two instruments directly aimed at assessing the risk of prison misconduct. One is the RASP (Risk Assessment Scale for Prison)[6] including a version for long-term prisoners, the RASP-Cap, developed in a sample of 136 incarcerated capital murder offenders in Texas.[7] The second instrument is a hybrid assessment system, developed by Makarios and Latessa,[8] consisting of a reduced classification instrument, a case management screen, and a full case management instrument.

In lack of well known validated instruments to assess prison misconduct, general risk assessment instruments are commonly used. One such instrument is the VRAG (Violence Risk Appraisal Guide).[9] From a study of 473

1. Campbell, M. A., French, S. and Gendreau, P. (2009) The prediction of violence in adult offenders. A meta-analytic comparison of instruments and methods of assessment. *Criminal Justice and Behavior*, 36, 567–590.
2. Singh, J. P., Grann, M. and Fazel, S. (2011) A comparative study of violence risk assessment tools: A systematic review and meta-regression analysis of 68 studies involving 25,980 participants. *Clinical Psychology Review, 31*, 499–513.
3. Haggård-Grann, U. (2007) Assessing violence risk: A review and clinical recommendations. *Journal of Counseling and Development, 85*, 295–302.
4. Bonta, A. (2002) Offender risk assessment: Guidelines for selection and use. *Criminal Justice and Behavior, 29*, 355–379.
5. Kroner, D. G. and Mills, J. F. (2001) The accuracy of five risk appraisal instruments in prediction institutional misconduct and new convictions. *Criminal Justice and Behavior, 28*, 471–489.
6. Cunningham, M. D., Sorensen, J. R. and Reidy, T. J. (2005) An actuarial model for assessment of prison violence risk among maximum security inmates. *Assessment*, 12, 40–49.
7. Cunningham, M. D. and Sorensen, J. R. (2007) Predictive factors for violent misconduct in close custody. *The Prison Journal, 87*, 241–253.
8. Makarios, M. and Latessa, E. J. (2013) Developing a risk and needs assessment instrument for prison inmates. The issue of outcome. *Criminal Justice and Behavior, 40*, 1449–1471.
9. Quinsey, V. L., Harris, G. T., Rice, M. E. and Cormier, C. A. (1998) *Violent offenders: Appraising and managing risk*. Washington DC: American Psychological Association.

male and female inmates in Washington DC,[10] it was reported that the VRAG predicted misconduct among male but not among female inmates. In Swiss study,[11] VRAG was used to predict misconduct in a sample of 106 violent offenders and sex offenders. The results indicated that the VRAG predicted misconduct among the sex offenders only. Other risk assessment instruments reported in studies of the prediction of prison misconduct are the PPI (Psychopathic Personality Inventory),[12,13] the PICTS (The Psychological Inventory of Criminal Thinking Style),[14,15] and the PAI (The Personality Assessment Inventory).[16, 17]

There are also studies in which instruments are compared. In a meta-analysis of misconduct and reconviction comprising 88 studies, Campbell and colleagues[1] compared the predictive validity of the HCR-20 (Historical, Clinical, Risk-20),[18] the LSI-R (Level of Services Inventory-Revised),[19] the PCL-R (Psychopathy Checklist-Revised),[20] and the VRAG. The authors concluded that standardized instruments, based on statistically derived risk factors, had the best predictive validity for prison misconduct. Kroner and Mills[5] conducted a comparative study of five different instruments among 97 inmates in Ontario, Canada. The instruments were the LSI-R, the HCR-20, the PCL-R, the VRAG and the LCSF (Lifestyle Criminality Screening Form).[21] The results showed similar predictive validity for all of the instruments.

It could be expected that prisoners serving long-term sentences would cause more trouble in prison as compared to short-term sentenced inmates.

Lifetime prisoners

The literature on lifetime prisoners and prison misconduct is scarce. It could be expected that prisoners serving long-term sentences would cause more trouble in prison as compared to short-term sentenced inmates ('nothing to lose').[22] However, research does not support this assumption and even indicates that the reverse may be the case. Cunningham and Sorensen reported from of a study of inmates sentenced to life without parole (n = 1897) and long-term inmates serving at least 10 years (n = 7147) in Florida that the likelihood and pattern of prison misconduct were similar between subsamples.[22] Morris and colleagues[23] reviewed criminal files of capital inmates in Texas with sentences that differed in number of years before becoming eligible for parole (15 years, n = 71; 35-40 years, n = 329). The prisoners with longer sentences before parole were found to be less likely to engage in serious misconduct than those with shorter sentences before parole. Potential differences between long-term and short-term inmates that may have an impact on misconduct, e.g., older age, maturation, personality

10. Hastings, M. E., Krishnan, S., Tangney, J. P. and Stuewig, J. (2011) Predictive and incremental validity of the Violence Risk Appraisal Guide scores with male and female jail inmates. *Psychological Assessment, 23*, 174–183.
11. Endrass, J., Rossegger, A., Frischknecht, A., Noll, T. and Urbaniok, F. (2008) Using the Violence Risk Appraisal Guide (VRAG) to predict in-prison aggressive behaviour in a Swiss offender population. *International Journal of Offender Therapy and Comparative Criminology, 52*, 81–89.
12. Lilienfeld S. O. and Andrews, B. P. (1996) Development and preliminary validation of a self-report measure of psychopathic personality traits in noncriminal population. *Journal of Personality Assessment, 66*, 488–524.
13. Edens, J. F., Lilienfeld, S. O., Poythress, N. G., Patrick, C. J. and Test, A. (2008) Further evidence of the divergent correlates of the psychopathic personality inventory factors: Predictions of institutional misconduct among male prisoners. *Psychological Assessment, 20*, 86–91.
14. Walters, G. D. (1995) The Psychological Inventory of Criminal Thinking Styles: Part I. Reliability and preliminary validity. *Criminal Justice and Behavior, 27*, 307–325.
15. Walters, G. D. and Schlauch, C. (2008) The Psychological Inventory of Criminal Thinking Styles and Level of Service Inventory-Revised. *Law and Human Behavior, 32*, 454–462.
16. Morey, L. C. (1991) *Personality Assessment Inventory: Professional manual*. Odessa, FL: Psychological Assessment Resources.
17. Newberry, M. and Shuker, R. (2012) Personality assessment inventory (PAI) profiles of offenders and their relationship to institutional misconduct and risk of reconviction. *Journal of Personality Assessment, 94*, 586–592.
18. Webster, C. D., Eaves, D., Douglas, K. S. and Wintrup, A. (1995) The HCR-20 scheme: The assessment of dangerousness and risk. Vancouver, Canada: Simon Fraser University and British Columbia Forenic Psychiatric Services Commission.
19. Andrews, D. A. and Bonta, J. (1995) *Level of Service Inventory-Revised*. Toronto, Canada: Multi-Health Systems.
20. Hare, R. D. (2003) *The Hare Psychopathy Checklist — Revised. (2nd ed.): Manual*. Toronto, Canada: Multi-Health Systems.
21. Walters, G. D., White, T. W. and Denney, D. (1991) The Lifestyle Criminality Screening Form: Preliminary data. *Criminal Justice and Behaviour, 18,* 406–418.
22. Cunningham, M. D. and Sorensen, J. R. (2006) Nothing to lose? A comparative examination of prison misconduct rates among life-without-parole and other long-term high-security.
23. Morris, R. G., Longmire, D. R., Buffington-Vollum, J. and Vollum, S. (2010) Institutional misconduct and differential parole eligibility among capital inmates. *Criminal Justice and Behavior, 37*, 417–438.

change and a preference for solo but pro-social activities like reading, have been discussed among researchers.[24, 25] This is in line with reports from staff from lifetime prison wards in Sweden, pointing out that lifetime inmates are more likely to view the prison as their home in contrast to short-term inmates who rather see it as a temporary accommodation.[26]

The present study

The present study explores risk assessment of prison misconduct in a total cohort of lifetime prisoners in Sweden. Commonly used instruments for violence risk assessment of lifetime prisoners have until recently been the PCL-R, the HCR-20 and the VRAG, all of them validated over ten years ago in various Swedish settings PCL-R;[27, 28] VRAG;[29, 30] HCR-10.[29, 31, 32] The LSI-R has recently been translated into Swedish and is now used as a standard tool for risk assessments of lifetime prisoners applying for a fixed-term sentence.[33] It has not yet been validated in a Swedish setting.

Aim

The aim of the present study was to increase knowledge on risk assessment instruments of prison conduct among lifetime prisoners.

More specifically, in a total cohort of lifetime prisoners, sentenced between 1965 and 2007, the objective of the study was to examine the predictive validity of three commonly used risk assessment instruments, the LSI-R, the VRAG and the PCL-R, in the prediction of prison misconduct. A specific focus was devoted to the LSI-R, since the instrument has recently come into use in Sweden.

Method

Lifetime imprisonment in Sweden

The number of lifetime convictions in Sweden is low but has increased considerably during the past two decades. (1970–1989, $n = 59$; 1990–2009, $n = 119$.) [34, 35] In the year of 2014, the Swedish prison system held 144 lifetime prisoners.[36] In addition, the time served before release has been prolonged. During the 70s and 80s, petitions for mercy, made to the government, usually led to release from prison after seven years. However, practice changed and during the mid-90s incarcerations of lifetime prisoners usually lasted for 12–15 years. After changes in the Swedish legal system in 2006, applications of time-limited sentences are now made to the court and the shortest time possible to serve is 12 years. For those lifetime prisoners who have received court decisions between 2006 and 2012, the time of incarceration has varied from 12 to 30 years (detailed statistics not available).

Participants

All offenders convicted to lifetime imprisonment for murder or terrorism (leading to death) in Sweden between January 1965 and June 2007 were included into the study. Of those identified ($N = 248$), three individuals were excluded because of incomplete identification numbers and an additional three individuals were excluded because of missing criminal files. Hence, a total number of 242 lifetime prisoners were included into the study. At the time of the end of study (April 31st, 2009), 162 of the participants (66.9 per cent) were still serving their prison sentence while 55 participants (22.7 per cent) had been released from prison by petition for mercy or by

24. Manchak, S. M., Skeem, J. L. and Douglas, K. S. (2008) Utility of the Revised Level of Service Inventory (LSI-R) in predicting recidivism after long-term incarceration. *Law, Human and Behavior, 32*, 477–488.
25. Toch, H. (2010) 'I am not now who I used to be then'. Risk assessment and the maturation of long-term prison inmates. *The Prison Journal, 90*, 4–11.
26. Personal communication with staff at the Swedish Prison and Probation Service 2007-05-23.
27. Grann, M., Långström, N., Tengström, A. and Kullgren, G. (1999) Psychopathy (PCL-R) predicts violent recidivism among criminal offender with personality disorders in Sweden. *Law and Human Behavior, 23*, 205–217.
28. Långström, N. and Grann, M. (2002) Psychopathy and violent recidivism among young criminal offenders. *Acta Psychiatrica Scandinavica Supplement, 412*, 86–92.
29. Grann, M., Belfrage, H. and Tengström, A. (2000) Actuarial assessment of risk for violence: Predictive validity of the VRAG and the historical part of the HCR-20. *Criminal Justice and Behavior, 27*, 97–114.
30. Grann, M., Långström, N., Tengström, A. and Kullgren, G. (1999) Psychopathy (PCL-R) predicts violent recidivism among criminal offender with personality disorders in Sweden. *Law and Human Behavior, 23*, 205–217.
31. Belfrage, H., Fransson, R. and Strand, S. (2000) Prediction of violence using the HCR-20: a prospective study in two maximum-security correctional institutions. *The Journal of Forensic Psychiatry, 11*, 167–175.
32. Dernevik, M. (1998) Preliminary findings on reliability and validity of the historical-clinical-risk assessment in a forensic psychiatric setting. *Psychology, Crime and Law, 4*, 127–137.
33. National Board of Forensic Medicine. Website. http://www.rmv.se/index.php?id=213 Updated May 29, 2015. Accessed November 27, 2015.
34. Swedish Government Official Reports, SOU (2002). *Betänkande från Utredningen om frigivningsprövning av livstidsdömda* [Commission report on the probational release of lifetime offenders]. Justiedepartmentent (Swedish Ministry of Justice): 2002:26. Retrieved December 3, 2013, from http://www.regeringen.se/sb/d/136/a/482.
35. National Prison and Probation Services. Statistics of convictions to lifetime imprisonment l in Sweden 2003–2013, http://www.kriminalvarden.se/Statistik/Livstidsdomda/Statistik-over-livstidsdomda. Accessed April 16, 2013.
36. National Prison and Probation Services. Statistics of convictions to lifetime imprisonment in Sweden. http://www.kriminalvarden.se/forskning-och-statistik/statistik-och-fakta/kriminalvardens-pafoljder/#livstidsstraff. Accessed November 27, 2015.

a court decision. The remaining 25 participants (10.3 per cent) had either been transferred to their home countries to serve their sentences (*n* = 16, 6.6 per cent), were dead (*n* = 6, 2.5 per cent), had absconded (*n* = 2, 0.8 per cent) or had the sentence converted to forensic psychiatric care (*n* = 1, 0.4 per cent).

As shown in Table 1, almost all of the participants were male and in their mid thirties when committing the index offence. Half of the participants were of Swedish origin. They had low educational levels and lacked work stability. More than half of the participants suffered from one or several mental health problems. One fifth of the participants had a PCL-R score of 27 or more (a recommended cut-off score for retrospective assessments of psychopathy based on files).[30, 37] Worth noting is that those participants convicted of more than one murder (*n* = 24, 9.9 per cent) had a median PCL-R score of 30.2. The vast majority of the participants had been subjects of correctional sanctioning prior to index offence.

Table 1. Demographic description of the total cohort of lifetime prisoners 1965 — 2007 (*N* = 242).

Sex	
Male	236 (97.5 per cent)
Female	6 (2.5 per cent)
Nationality	
Swedish	121 (50.0 per cent)
Nordic (excluding Swedish)	34 (14.0 per cent)
European (excluding Nordic)	46 (19.0 per cent)
Others (excluding European)	41 (16.9 per cent)
Education/employment	
Number of completed school years	*M* = 9.7 (SD = 2.5)
Not completed compulsory school	64 (26.4 per cent)
Never employed for a full year	97 (40.1 per cent)
Offence history	
Age at index offence	*M* = 35.5 (SD = 9.5)
Index offence, number of victims	
One victim	218 (90.1 per cent)
More than one victim (range 2–7)	24 (9.9 per cent)
Prior convictions	
At least one prior conviction	173 (71.5 per cent)
Three or more convictions	124 (51.2 per cent)
At least one prior imprisonment	139 (57.4 per cent)
Mental health	
Alcohol misuse	112 (46.3 per cent)
Drug abuse	78 (32.2 per cent)
Personality disorder	135 (55.8 per cent)
Psychopathy (PCL-R ≥ 27)	48 (19.8 per cent)
Major mental disorder	17 (7.0 per cent)

Procedure

The study was conducted through retrospective risk assessments using data from the correctional records of the participants, conducted during July 2007 to December 2008. The raters were the first author and three research assistants all with long clinical experience from forensic psychiatry and authorized raters of the LSI-R and the PCL-R. The data used for the study included all information from the time of the index crime until a few months after the sentence, e.g., police reports, evaluations by the probation services, court hearings, initial assessments within the prison system and, in those cases available, forensic psychiatric evaluations (FPE). According to the Swedish legislation, the criminal court can refer the defendant to a forensic psychiatric evaluation (FPE) if it is suspected that the offence was committed under the influence of a severe mental disorder. A minor FPE is based on files and an interview by a forensic psychiatrist. A major FPE is usually made during four weeks of observations and interviews by a multidisciplinary forensic psychiatric team addressing medical, psychiatric, psychological and social aspects of the individual and the offence. Almost all of the participants, *n* = 234 (97 per cent) had undergone either a minor or a major FPE.

Risk assessment instruments

The risk assessment instruments used in the study were the LSI-R, the PCL-R and the VRAG. The Level of Service Inventory — Revised (LSI-R)[19] is a checklist based on the risk-need-responsivity model of correctional assessment and crime prevention.[38] It has been widely studied in offender populations[39] and has become commonly used as one of the standard instruments in the US.[24] The LSI-R comprises 54 items in 10 subscales: criminal history, education/employment, financial, family/marital, accommodation, leisure/recreation, companions, alcohol/drug problem, emotional/personal,

37. Dåderman, A. M. and Kristiansson, M. (2003) Degree of psychopathy: Implications for treatment in male juvenile delinquents. *International Journal of Law and Psychiatry, 26*, 301–315.
38. Andrews, D. V., Bonta, J. and Wormith, J. S. (2011) The Risk-Need-Responsivity (RNR) Model: Does adding the Good Lives Model contribute to effective crime prevention? *Criminal Justice and Behavior, 38*, 735–755.
39. Andrews, D. V. and Bonta, J. (2003) *The level of service inventory-revised U.S. norms manual supplement*. Toronto, Canada: Multi-Health Systems.

and attitudes/orientation. Two thirds of the items in the LSI-R are based on dynamic factors and are tightly correlated to potential treatment areas. The inter-rater reliability for trained raters suggests acceptable levels (*ICC*=0.80-0.96).[5, 19] A professional override is a part of the LSI approach.[38]

The Psychopathy Checklist — Revised (PCL-R)[20] was developed to assess the degree of psychopathic personality traits in an individual. Since psychopathy has been shown to be one of the strongest individual predictors of violence and violent recidivism among adults,[40, 41] it has been widely used for the assessment of future criminal acts.[42] The checklist consists of 20 items, each of them scored from 0 to 2. Factor analyses have yielded two-, three- and four-factor models.[20,43,44] Inter-rater agreement of the PCL-R shows ICC in the range 0.60[45] to 0.88.[20]

The Violence Risk Appraisal Guide (VRAG)[46] is an oft-used actuarial instrument aimed at the assessment of violence risk. [11, 47] The instrument is constructed so that no clinical training is required, except for the rating of psychopathy using the PCL-R. The instrument comprises 12 items. Each variable is weighted, with psychopathy having the greatest weight. All items are stable, i.e., they are not likely to change over time. The scores are added into a total score ranging from -26 to 38. Individuals are assigned to one of nine risk categories, ranging from 1 (lowest risk) to 9 (highest risk), according to their total score. The inter-rater reliability (ICC) for the VRAG has shown results in the high range 0.92,[48] even excellent (r = 1.0) when using risk categories.[49] However, it has been demonstrated that some variables show lower inter-rater reliability in studies based on retrospective and file-based information, namely childhood variables such as elementary school maladjustment and separation from parents.[50]

The rational not to use the HCR-20 in the study was based on its psychiatric focus and the difficulty to extract information from the correctional records on the clinical variables. To prevent any bias, all risk assessments were rated blind to outcome in terms of prison misconduct.

Measures of outcome

Information on prison misconduct was acquired from the correctional records. Two types of misconduct were recorded in the files: minor and major misconduct. Minor misconduct involved improper dressing, non-compliance with common rules, invalid absence from work and similar. The choice for the present study was not to include minor misconduct. Major misconduct comprised threat and assault towards staff/prisoners, possession of weapon, incitement of a riot, drug/alcohol use, refusal of urinalysis, and disobedience of a direct order. Major misconduct was coded into two categories, high frequency of misconduct and severe misconduct. High frequency of misconduct was defined as 10 or more incidents of major misconduct. Severe misconduct was defined as at least one incident of threat or violence. The follow-up time within the prison period differed between participants (M = 94.0 months, SD = 79.4, range 1 — 348).

Statistical analyses

The inter-rater agreement and concurrent validity

The inter-rater agreement between the four raters was calculated through the two-way mixed Analysis of Variance (ANOVA), absolute agreement random model type Intra Class Correlation (ICC), for the continuous variables[51] in a random subsample (n=22) (Table 2). Pearson correlation coefficients, r, were reported for the purpose of establishing

40. Salekin, R. T., Rogers, R. and Sewill, K. W. (1996) A review and meta-analysis of the Psychopathy Checklist and Psychopathy Checklist — Revised: Predictive validity of dangerousness. *Clinical Psychology: Science and Practice, 3*, 203–215.
41. Steadman, H. J., Silver, E., Monahan, J., Appelbaum, P. S., Robbins, P. C., Mulvey, E. P., et al. (2000) A classification tree approach to the development of actuarial violence risk assessment tools. *Law and Human Behavior, 24*, 83–100.
42. Archer, R. P., Buffington-Vollum, J. K., Stredny, V. R. and Handel, R. W. (2006) A survey of psychological test use patterns among forensic psychologists. *Journal of Personality Assessment, 87*, 84–94.
43. Cooke, D. J. and Michie, C. (2001) Refining the construct of psychopathy: Towards a hierarchical model. *Psychological Assessment, 13*, 171–188.
44. Hare, R. D. (1991) *The Hare Psychopathy Checklist — Revised: Manual*. Toronto, Canada: Multi-Health Systems.
45. Miller, C. S., Kimonis, E. R., Otto, R. K., Kline, S. M. and Wasserman, A. L. (2012) Reliability of risk assessment measures used in sexually violent predator proceedings. *Psychological Assessment, 24*, 944–953.
46. Harris, G. T., Rice, M. E. and Quinsey, V. L. (1993) Violent recidivism of mentally disordered offenders: The development of a statistical prediction instrument. *Criminal Justice and Behavior, 20*, 315–335.
47. Harris, G. T., Rice, M. E. and Cormier, C. A. (2002) Prospective replication of the Violence Appraisal Guide in predicting violent recidivism among forensic patients. *Law and Human Behavior, 26*, 377–394.
48. Douglas, K. S., Yeomans, M. and Boer, D. P. (2005) Comparative validity analysis of multiple measures of violence risk in a sample of criminal offenders. *Criminal Justice and Behavior, 32*, 479–510.
49. Lofthouse, R. E., Lindsay, W. R., Totsika, V., Hastings, R. P., Boer, D. P. and Haaven, J. L. (2013) Prospective dynamic assessment of risk of sexual reoffending in individuals with an intellectual disability and a history of sexual offending behaviour. *Journal of Applied Research in Intellectual Disabilities, 26*, 394–403.
50. Långström, N., Grann, M., Tengström, A., Lindholm, N., Woodhouse, A. and Kullgren, G. (1999) Extracting data in file-based forensic psychiatric research: Some methodological considerations. *Nordic Journal of Psychiatry, 53*, 61–67.
51. Dunn, G. and Everitt, B. S. (2004) *An Introduction to Mathematical Taxonomy*. Courier Dover Publications.

concurrent validity between the risk assessment instruments. To briefly explain the analysis methods, an ANOVA test is used to test differences between means when there are more than two groups involved in the analysis. ICC measures the level of inter-rater agreement. Pearson correlation coefficients, finally, shows the degree of relationship between two variables.

Predictive validity

Receiver Operating Characteristics (ROC)[52] was used to analyse the predictive validity of the risk assessment instruments. ROC is commonly viewed as being fairly stable and independent of base rates and selection ratios when compared to other prediction methods.[53] A ROC curve can be used to explore how the specificity is affected as the sensitivity increases with the area under the curve (AUC) as an estimate of the overall accuracy of a certain measure for the prediction of a dichotomous outcome. An AUC value of .50 means no accuracy and 1.0 means perfect accuracy.[52] There are no fixed interpretations of the AUC estimates[54] but an area of .75 and above has been suggested as 'large',[55] although this has been criticised of being overly optimistically interpreted.[56]

All statistical analyses were performed using the statistical software package SPSS version 19.0.

Results

Incidents of misconduct

One third of the participants (*n* = 77, 32 per cent) had been reported for 10 or more incidents of major misconduct at the time of release or at the end of the study, whichever was first. Four out of ten of the participants (*n* = 100; 41 per cent) had been reported for a least one threat and one third of them (*n* = 82; 34 per cent) had been reported for at least one incident of violence. Almost half of the participants (*n* = 111; 46 per cent) had been reported for either.

Descriptive statistics of the risk assessment instruments

As shown in Table 2, the retrospective ratings of the participants with the LSI-R, the PCL-R, and the VRAG and showed a large variation among participants from very low to very high risk of misconduct. The inter-rater reliability between the raters indicated that the reliability was highest for the LSI-R (0.93) and lowest for the VRAG (0.66) (see Table 2). The LSI subscales of family/marriage and emotional/personal had the lowest inter-rater reliability scores among the LSI subscales. As expected, the inter-rater agreement was lower for Factor 1 of the PCL-R as compared to Factor 2.

Table 2. Descriptive statistics and inter-rater agreement of the risk assessment instruments LSI-R, PCL-R and VRAG.

	LSI-R	PCL-R	VRAG
N	240	241	240
M (range)	23.4 (2-46)	17.1 (0-37)	0.13 (-22-34)
SD	11.7	9.7	12.0
Inter-rater reliability	**Total** .93 (.05 — .98)	**Total** .80 (0.34 — 0.93)	.66 (-.02 — .87)
(*n* = 22) ICC (95 per centCI)	Criminal history .94 (.87 — .98)	Factor 1 .64 (.16 — .85)	
	Education/Employment .79 (.45 — .92)	Factor 2 .85 (.64 — .94)	
	Financial .82 (.25 — .94)		
	Family/Marital .23 (-.73 — .67)		
	Accommodation .88 (.71 — .95)		
	Leisure/Recreation .90 (.77 — .96)		
	Companions .77 (.44 — .90)		
	Alcohol/Drug problem .90 (.75 — .96)		
	Emotional/Personal .14 (-.63 — .60)		
	Attitudes/Orientation .76 (.44 — .90)		
per cent Complete cases	99.2	99.6	99.2

52. Hanley, J. A. and McNeil, B. J. (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology, 143*, 29–36.
53. Singh, J. P., Desmarais, S. L. and van Dorn, R. A. (2013) Measurement of predictive validity in violence risk assessment studies: A second-order systematic review. *Behavioral Sciences and the Law, 31*, 55–73.
54. Andrews, D.V., Bonta, J., Wormith, J. S., Guzzo, L., Brews, A., Rettiger, J., et al. (2011) Sources of variability in estimates of predictive validity: A specification with Level of Service general risk and need. *Criminal Justice and Behavior, 28*, 413–4.
55. Dolan, M and Doyle, M. (2000) Violence risk prediction: Clinical an actuarial measures and the role of the Psychopathy Checklist. *British Journal of Psychiatry, 177*, 303–311.
56. Sjöstedt, G. and Grann, M. (2002) Risk assessment: What is being predicted by actuarial prediction instruments? *International Journal of Forensic Mental Health, 1*, 179–183.

Pearson correlations showed positive and significant (p <0.01, two-tailed) correlations between all instruments (PCL-R — LSI-R, r = 0.69; PCL-R — VRAG, r = 0.79; LSI-R — VRAG, r = 0.75). Subanalyses of the two-factor model of the PCL-R displayed smaller correlations between PCL-R Factor 1 and the other two instruments as compared to PCL-R Factor 2 and the two instruments (PCL-R Factor 1 — LSI-R, r = 0.38; PCL-R Factor 2 — LSI-R, r = 0.77; PCL-R Factor 1 — VRAG, r = 0.51; PCL-R Factor 2 — VRAG, r = 0.81).

The predictive validity of the instruments

As shown in Table 3, the predictive validity of the three instruments was very similar. All three instruments, LSI-R, PCL-R, and the VRAG made slightly better predictions of high frequency of misconduct (.71; .70; .70, respectively) as compared to the predictions of severe misconduct (.65; .69; .67, respectively). An unexpected finding was that the LSI-R subscale of attitude/orientation, consisting of items on attitudes to criminality, reflections upon the harm inflicted to possible victims, and orientation towards a conventional life, was the single best predictor of both high frequency of misconduct and severe misconduct. The least accurate LSI-R subscales in their predictions of prison misconduct were family/marriage and emotional/personal.

Table 3. The predictive validity of the LSI-R, PCL-R and VRAG on high frequency of misconduct and severe misconduct during imprisonment.

Instrument	High frequency of misconduct (≥10) AUC of ROC (95 per centCI)	Severe misconduct (threat and violence) AUC of ROC (95 per centCI)
LSI-R (total)	.71 (.64 — .78)	.65 (.58 — .72)
Criminal History	.68 (.61 — .76)	.61 (.54 — .69)
Education/Employment	.65 (.57 — .72)	.65 (.57 — .72)
Finances	.63 (.55 — .70)	.54 (.46 — .61)
Family/Marriage	.48 (.40 — .56)	.55 (.47 — .62)
Accommodations	.58 (.50 — .66)	.58 (.50 — .65)
Leisure/Recreation	.61 (.53 — .69)	.61 (.54 — .69)
Companions	.70 (.62 — .77)	.66 (.59 — .73)
Alcohol/Drugs	.70 (.62 — .77)	.58 (.50 — .65)
Emotional/Personal	.51 (.43 — .59)	.51 (.43 — .58)
Attitude/Orientation	.76 (.69 — .82)	.72 (.66 — .79)
PCL-R Total	.70 (.63 — .77)	.69 (.62 — .76)
Factor 1	.61 (.54 — .69)	.66 (.58 — .73)
Factor 2	.71 (.64 — .78)	.66 (.59 — .73)
VRAG	.70 (.63 — .77)	.67 (.60 — .74)

Note. LSI-R=Level of service Inventory-Revised; PCL-R=Psychopathy Checklist-Revised; VRAG=Violence Risk Appraisal Guide.

AUC of ROC=Area under the curve of the receiver operating characteristics.

Discussion

Violence risk assessment is an important aspect of decision making within the correctional service. The results of the present study showed that major prison misconduct was fairly common among lifetime prisoners in Sweden. One third of the participants had been reported for 10 or more incidents of major misconduct and four out of ten of the participants had been reported for either a threat or an incident of violence. The large proportion of lifetime prisoners involved in prison misconduct is similar to results from studies on lifetime prisoners in the US.[22, 23]

The main finding of the study was that the three risk assessment instruments used in the study were similar as to their predictive accuracy. The predictive validity of high-frequency misconduct was moderately successful, but decreased with severe misconduct as the outcome measure. When interpreting the results it is important to keep in mind the basis of predictive research findings. The AUC of ROC in our setting reflects the likelihood that the risk score of a randomly chosen misconducting prisoner is higher than that of a randomly chosen non-misconducting prisoner. The results are not unexpected, given that the content of the instruments overlap to some extent. As an example, all three instruments include items on criminal history and personality related variables. The VRAG and the PCL-R have a similar item on early-onset behavioural problems, and the VRAG and the LSI-R share an item on substance misuse.

A note of caution should be introduced. While an AUC of .70 may seem impressive, it is a statistical measure based on the analyses of group data. Predictions of the probability that a proportion of individuals from a group may show certain behaviors are difficult, but not impossible to make, and the larger the group, the more accurate the prediction. Once the prediction is to be made for one single individual, the uncertainty of the prediction increases substantially. This has been extensively discussed within the scientific literature,[57, 58] and will not be further elaborated here.

With this note of caution and with the similar predictive accuracy of the risk assessment instruments explored in the current study, there may be further considerations to make when choosing a risk assessment instrument for lifetime prisoners. One such consideration may include the possibility of using the risk assessment as a guide for treatment and risk management. A lifetime sentence offers sufficient time to undergo interventions targeting the individual's criminogenic needs. This would require a theory-based instrument with a broad focus, including dynamic and changeable factors, such as the

57. Hart, S. D., Michie, C. and Cooke, D. J. (2007). Precision of actuarial risk assessment instruments. Evaluation the 'margins of error' of group v. individual predictions of violence. *British Journal of Psychiatry, 190*, 60–65.
58. Hanson, R. K. and Howard, P. D. (2010). Individual confidence intervals do not inform decision-makers about the accuracy of risk assessment evaluations. *Law and Human Behavior, 34*, 275–281.

LSI-R. It has been argued that dynamic risk factors are more relevant for longer-term predictions of misconduct than they are for shorter-term predictions.[1] The VRAG includes a diverse set of risk factors, but they are unchangeable and thus give little guidance for the practitioner on how to lower the risk. The PCL-R, on the other hand, includes a number of dynamic risk factors, but has, in contrast to the other two instruments, a narrow focus on personality traits.

We have proposed that the LSI-R may be a valid instrument not only for risk assessment but also as a starting point for risk management and treatment. Interestingly, treatment to reduce prison misconduct may also have effects on recidivism after release. In a meta-analysis by French and Gendreau[59] it was concluded that prison-based behavioural programs produce large reductions in misconduct that may carry over into reductions in recidivism in the community.

Strengths and limitations

The participants of our study constituted a total national cohort of lifetime prisoners, sentenced during a time period of more than 40 years. They were retrospectively assessed with well-validated risk assessment instruments. The information used was generally of high quality. The minor and major forensic psychiatric evaluations and the evaluations by the probation services are standardized and detailed. The raters of the study were forensic psychiatric social workers with long clinical experience from forensic psychiatry. Although the number of participants was limited, we could compare the predictive validity of the instruments (with fairly robust results), based on the high rates of outcome.

There were also some limitations. The quality of the information acquired from retrospective ratings of archive information may be questioned. Even though retrospective and file based assessments with different risk instrument have been proven to be useful for data extraction,[60] some items in the risk assessment instruments may not be possible to assess due to poor information quality.[50]

There may also be limitations related to the inter-rater reliability. The higher agreement for the LSI-R may be due to the more recent training provided to the raters. On the other hand, lower inter-rater agreements for the PCL-R, specifically, have been reported elsewhere,[45, 61] and can probably be explained by the fact that the rater need to make inferences regarding behavioural styles and personality characteristics. Additionally, it is important to remember that the differences in inter-rater reliability may further affect the validity of the instruments. For example, in this study we found that the two subcomponents in LSI-R showing the lowest inter-rater scores also had the lowest predictive validity. It is possible that the predictive validity of the VRAG would increase with higher inter-rater reliability. A problem with low reliability for PCL-R Factor 1 scores with retrospective file-based information has also been discussed.[50] The ROC statistics cannot compensate for low reliability of individual items.[62] To reduce problems with low inter-rater reliability due to poor quality of information, future studies within the Swedish Prison and Probation Services should preferably be prospective in their design.

A specific limitation refers to the areas of interest of the instruments used. LSI-R, VRAG and PCL-R are all developed with the individual prisoner at focus, thus ignoring the potential influence of the structural features of the institution such as social density, the existence of prison gangs, inmate-to-officer ratios, and security levels, factors that may also contribute to prison misconduct.[63, 64]

Caution is also warranted when considering the generalisability of these findings. The targeted population is a highly selective and a relatively small offender group.

Conclusions

This study was the first to investigate prison misconduct among lifetime prisoners in Sweden and also the first to validate the LSI-R, the VRAG and the PCL-R in the prediction of prison misconduct. The predictive validity between the risk assessment instruments was found to be similar. We therefore argue that additional aspects of the instruments should be considered. Such an aspect is a high inter-rater agreement. Another aspect is that the instrument should include dynamic and changeable factors. Finally, an important aspect is that the instrument should include guidelines to treatment and risk management. We strongly suggest that all of those factors should be considered in the choice of risk assessment instruments in prisons and other settings. From the results of the current study we advocate the use of LSI-R in the prediction of prison misconduct among lifetime prisoners in Sweden.

59. French, S. A. and Gendreau, P. (2006) Reducing prison misconducts. *Criminal Justice and Behavior, 33*, 185–218.
60. Grann, M., Långström, N., Tengström, A. and Stålenheim, E. G. (1998) The reliability of file-based retrospective rating of psychopathy with the PCL-R. *Journal of Personality Assessment, 70*, 416–426.
61. Edens, J. F., Boccaccini, M. T. and Johnson, D. W. (2010) Inter-rater reliability of the PCL-R total and factor scores among psychopathic sex offenders: Are personality features more prone too disagreement than behavioural features? *Behavioral Sciences and the Law, 28*, 106–119.
62. Harris, G. T. and Rice, M. E. (2003) Actuarial assessment of risk among sex offenders. *Annuals of the New York Academy of Sciences, 989*, 198–210.
63. Arbach-Lucioni, K., Martinez-García, M., and Andrés-Puejo, A. (2012) Risk factors for violent behaviour in prison inmates: A cross-cultural contribution. *Criminal Justice and Behavior, 39*, 1219–1239.
64. Griffin, M. L. and Hepburn, J. R. (2013) Inmate misconduct and the institutional capacity for control. *Criminal Justice and Behavior, 40*, 270–288.